Speech Power and Its Measurement [1]

By L. J. SIVIAN

The paper is chiefly concerned with the important speech power quantities—frequency spectra, distributions of instantaneous, average, syllabic and peak amplitudes, etc.—as they obtain in actual speech for a large range of voices, talking levels, and subject matters. The analysis is not nearly so complete nor so fine-grained as that which, in principle, can be derived from oscillographic records of individual speech sounds. Its advantage is in the speed with which data can be secured, under widely varying conditions and on a scale which warrants statistical conclusions. Some of the methods in use for measurements of this type are described. A "level analyzer" has been developed, primarily for the measurement of average and peak pressure amplitudes in speech and music, both as to magnitude and as to position in the frequency spectrum. Illustrative results are given for samples of speech, music and noise.

SPEECH sounds are so variable from one to another, from one individual to another, from one conversation to another, as to make it necessary to deal with speech power in a statistical manner. This is particularly true of engineering applications, as distinguished from studies in phonetics and voice dynamics. It is largely from the viewpoint of the former that the subject is here treated.

Speech power occurs in several states, e.g. acoustic, electric, magnetic, mechanical, optical, thermal, etc., but its measurement largely refers to speech in the acoustic or in the electric form. Acoustically, the simplest quantity to define is the instantaneous power transmitted through unit area tangent to the wave front. That power is $L = P \cdot U$, where P is the pressure and U the air particle velocity. With P expressed in bars (dynes per cm.^{2}), and U in cm./sec., L is given in ergs/cm.$^2 \times$ sec. We do not directly measure the product $P \cdot U$. No suitable means for doing that has been developed. In a progressive plane wave, or with good approximation in any progressive wave at sufficient distance from the source, P and U are in phase, and the expression for L simplifies down to $L = P^2/\rho c = U^2 \cdot \rho c$, where ρc is a constant equal to 41.5 mechanical ohms—the sound radiation resistance of air. Hence for the type of waves mentioned the wattmeter type of measurement (pressure $\times$ velocity, corresponding to voltage $\times$ current in the electrical case) may be replaced with the simpler measurement of pressure (voltage) or velocity (current) alone.

What are the acoustic voltmeters and ammeters that are capable

[1] Presented before Acoustical Society of America, May 11, 1929.

of measuring the instantaneous pressures and velocities respectively? First, as to acoustic ammeters: perhaps the two best known forms are the Rayleigh disk and the hot-wire microphone. The disk gives absolute values of U but its response is so slow that it can be used only to measure the effective values in comparatively sustained sound waves lasting, say one second or longer, or for the ballistic integration of shorter pulses. Hence it does not give a measure of the instantaneous velocities for even the slowest audio-frequency vibration. Furthermore, its use when exposed to the speaker in an open room, is rendered difficult by its susceptibility to spurious air currents. The only application of the disk to vocal power measurements which has come to my notice, is one by Prof. Zernov,[2] published in 1908. For sustained loud singing and shouting he found energy densities ranging from 0.3 to 2.0×10^{-4} ergs/cm.^{3}, at 2 meters distance from the singer. Assuming uniform distribution over a hemisphere of 2 m. radius, this gives a total power output of the voice of the order of 50,000 microwatts. At 2 m. distance reflections from the walls of the room materially raise the energy density as compared with that in a progressive wave.

Another acoustic ammeter is the hot-wire microphone. The resistance variations of the wire tend to follow the instantaneous values of the air particle velocity in the sound wave, but the sensitivity is a complicated function of the frequency, rapidly decreasing as the latter increases. Hence it does not give true oscillograms of speech sounds which in general include a frequency range of six or seven octaves.

Still another device in this general class is the glow-discharge microphone. Its response is determined largely by the amplitude of the air particle motion (E. Meyer, E.N.T., v. 6, 17-21, 1929). Hence, for constant sound intensity the response is inversely proportional to the frequency, roughly. No method for its absolute calibration has been proposed other than comparison with a microphone of known performance. The frequency response and the somewhat erratic behavior of the device in its present status, render it rather unsuitable for speech power studies.

Nearly all our information concerning speech power and the waveform of speech sounds has been obtained with acoustic voltmeters, i.e., with devices responding to pressure in the sound wave. To a first approximation, the ear belongs to this class. It includes the resonators which Helmholtz used in his vowel studies. Generally, the vital element in these devices is a diaphragm which vibrates with

[2] *Ann. der Phys.*, *26*, 94, 1908.

the alternating sound pressure and whose motion is converted (indirectly, as a rule) into a visual record. These acoustic voltmeters have gradually evolved from Scott's phonautograph and Koenig's manometric capsules and indicating flames of some 70 years ago to the present day technique. Their extensive use accounts for the fact that sound measurements frequently are expressed in terms of pressure rather than in terms of power. In many cases, when the relation between pressure and velocity is not known or too complex a function of frequency to be manageable,—the description of the sound wave must be confined to pressure values alone.

The first requisite for absolute measurements of speech pressures is a microphone which admits of an absolute electroacoustic calibration over the range of audio frequencies, and which has a substantially uniform sensitivity over the most important part of that range. The development of the condenser microphone and of the thermophone method for calibrating it, supplied this need. This fundamental contribution to the subject is due to E. C. Wente.[3] Using such a calibrated condenser microphone in conjunction with a vacuum tube amplifier and an oscillograph which were uniformly sensitive up to about 6000 p.p.s., I. B. Crandall and C. F. Sacia,[4] and I. B. Crandall [4] obtained oscillograms of the fundamental speech sounds in the English language. Essentially, these oscillograms give a picture of the instantaneous pressures throughout the duration of the sound, impressed on the microphone diaphragm at 2.5 cm. from the speaker's lips. A certain amount of similar work on German speech sounds has been published by Trendelenburg.[5] From the standpoint of phonetics these pressure amplitude oscillograms of individual sounds are a most comprehensive source of information. Their chief use has been in determining the frequency spectra of the fundamental speech sounds. Sacia [6] and Sacia and Beck [7] have used them to compute the mean and the peak powers of those sounds.

From the standpoint of engineering applications, speech power has a twofold interest: (*a*), there is the question of designing microphones, receivers, circuits, room acoustics, of controlling noise levels, etc., (*b*), the control of apparatus (chiefly adjustment of amplification) while it is handling speech from persons who acoustically are not controllable.

[3] *Phys. Rev.*, July 1917, April 1922, June 1922.
[4] *Bell Sys. Tech. Jour.*, April 1924 and Oct. 1925, resp.
[5] Wiss. Veroff. a. d. Siemens-Konzern, 1924 and 1925.
[6] *Bell Sys. Tech. Jour.*, Oct. 1925.
[7] *Bell Sys. Tech. Jour.*, July 1926.

In connection with (*a*) we can use data obtained under "laboratory" conditions, i.e. with the speakers, their voice levels, the talking speed, etc., suitably selected. Perhaps the simplest statistical characteristic

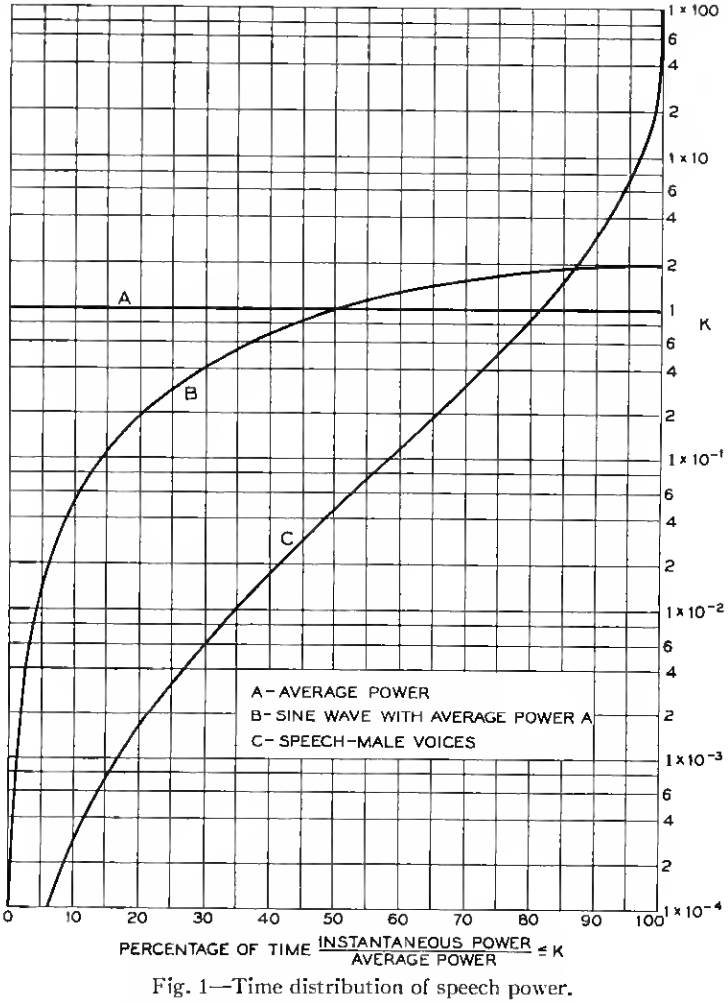

Fig. 1—Time distribution of speech power.

of speech is the average power, i.e. the ratio of the total energy in a large number of speech sounds divided by the total speech time. The quantity immediately measured is the pressure on the transmitter diaphragm. The corresponding power flow through unit area is computed. In doing so it is roughly assumed that the pressure on the diaphragm is twice as great as it would be in free air, which is

assuming total reflection by the transmitter. It is further assumed that the power flow is uniform over a hemisphere whose radius is the distance from the speaker's lips to the diaphragm. For this average power flow with "normally modulated" voices, Crandall and Mackenzie [8] found the value of 12.5 microwatts. More recently Sacia [6] gave 7.4 microwatts, including pauses between speech sounds, which

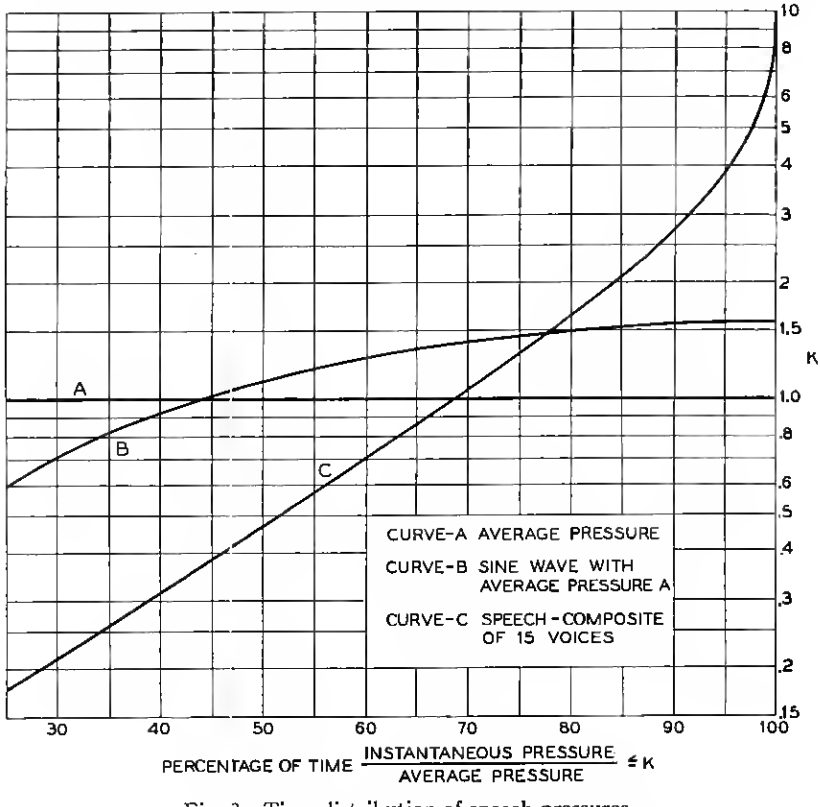

Fig. 2—Time distribution of speech pressures.

occupied about one-third of the total time. Some measurements made since then have led to substantially similar values, and so the value of 10 microwatts may be taken as fairly representative of a normal level. The corresponding average pressure on the diaphragm, at 5 cm. distance from the lips, was found to be about 5 bars. The averaging, of course, is for the absolute values of the pressures. Some notion of the level involved may be gained by noting th it is at

[8] *Phys. Rev.*, March 1922.

decidedly fatiguing to maintain a speech level 20 db higher for more than a few seconds.

It should be noted that the distributions of the instantaneous power and instantaneous pressure values in respect to their average values are quite different for speech from what they are in a sinusoidal wave. Fig. 1 and Fig. 2 show the extent of the difference for powers and pressures respectively. Clearly, speech as a whole is decidedly more "peaked" than a sinusoidal wave. Values of peak factors for

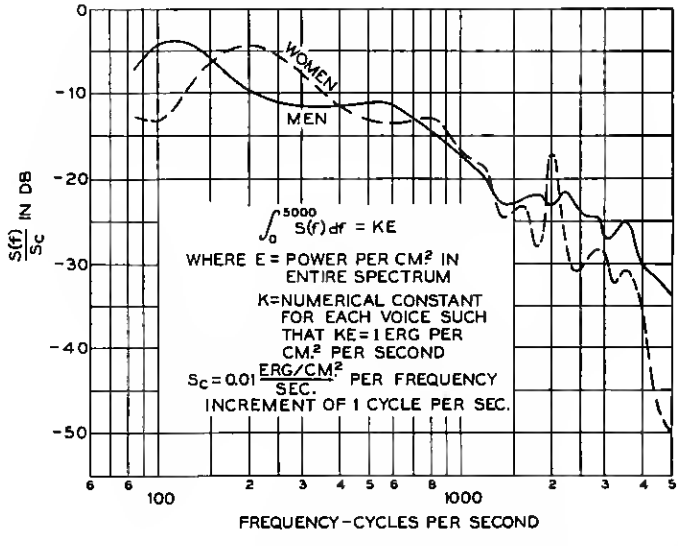

Fig. 3—Energy spectrum of connected speech-composite curves for 4 men and for 2 women.

individual speech sounds, i.e. the ratio of peak amplitude to mean effective amplitude, are given in Sacia's paper.[5]

A paramount characteristic of speech is the distribution of power and pressure in the frequency spectrum. That is so because in general the systems responding to or carrying speech (including the ear) have pronounced frequency characteristics. The same is true of noise sources interfering with speech. Here, as for total powers and pressures, the simplest quantity is the average spectrum for a large number of speech sounds. Crandall and Mackenzie[8] obtained the energy spectrum shown in Fig. 3, based on speech from six voices. They used a condenser microphone whose output was analyzed by a series of resonant circuits covering the range from 75 to 5000 p.p.s. Fig. 4 shows a recent determination in which the average pressure is given as a function of frequency. The apparatus

used to obtain the data will be described in detail elsewhere. Briefly, a series of band-pass filters are used to separate the frequency components. The output of any one filter goes into a rectilinear vacuum

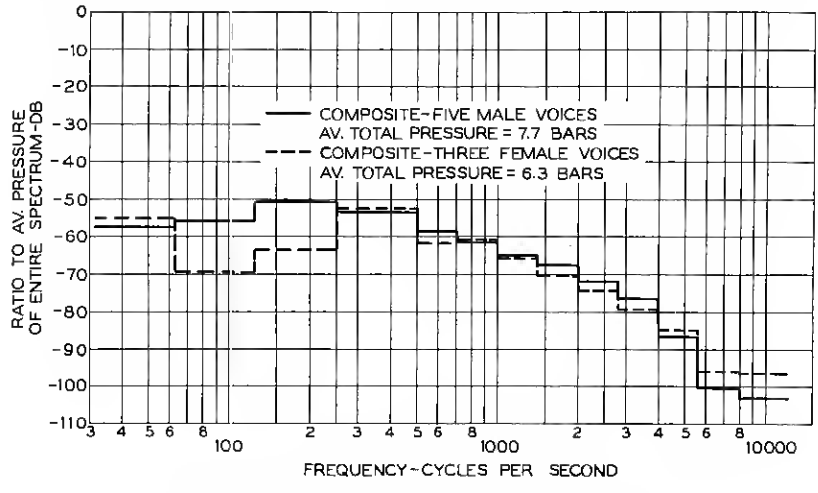

Fig. 4—Average speech pressures per frequency interval of 1 cycle per second—normal conversational voice. Distance 2″.

tube rectifier, and a fluxmeter integrates the rectified current over the duration of the speech. Simultaneously with this measurement the total spectrum also is rectified and integrated. The integration periods used were 15 seconds long. The pressure in any one band is

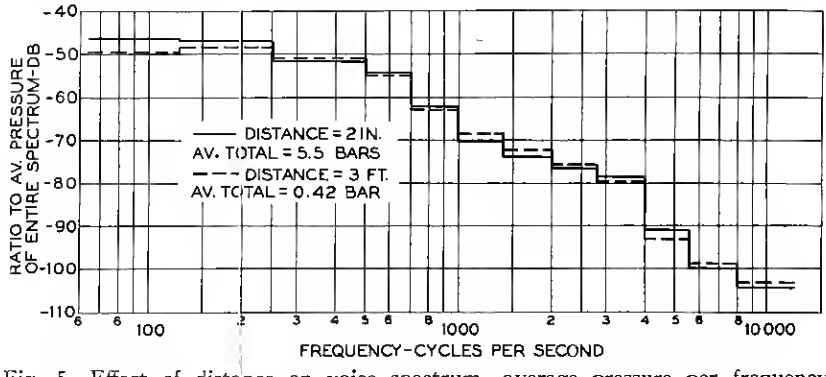

Fig. 5—Effect of distance on voice spectrum—average pressure per frequency interval of 1 cycle per second—normal speech-composite, three male voices.

given by the product: ($10^{0.05\alpha}$ × band width in cycles × total pressure), where α is the ordinate expressed in decibels (db). In using average spectra of this sort it is well to remember that they are

determined not only by the amplitudes in the individual speech sounds but equally by frequency of occurrence. Thus the fact that nearly all vowels for male voices at normal levels have a fundamental frequency between 80 and 150 p.p.s. tends to accentuate that region even though the fundamental amplitudes in individual speech sounds not be outstandingly large.

The data in Fig. 4 are for close talking conditions, 5 cm. from the lips to the diaphragm. Fig. 5 shows the effect of distance on the spectral distribution. The two distances are 5 cm. and 90 cm. respectively, in both cases the transmitter being set into a large fiber-board wall. The shapes of the two spectra are almost identical

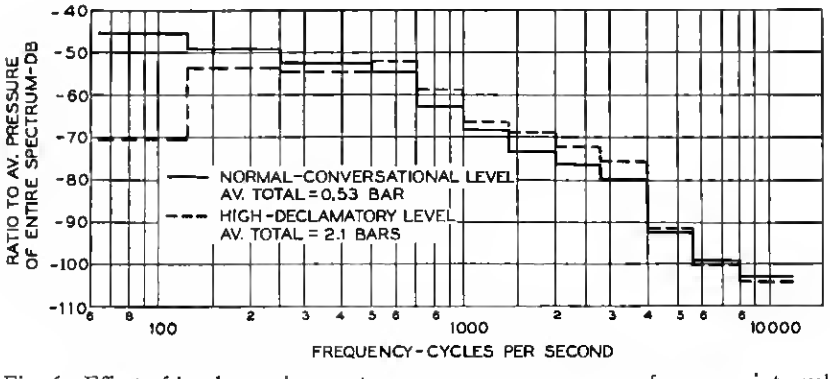

Fig. 6—Effect of level on voice spectrum—average pressures per frequency interval of 1 cycle per second—composite, three male voices, distance 3 ft.

indicating that, on the average, even at 5 cm. the condenser microphone does not greatly affect the voice as a sound generator. The largest difference between the two is in the lowest band, from 62 to 125 p.p.s., perhaps owing to the relatively low radiation efficiency of the voice at those frequencies.

The ratio of the distances is 18 : 1, that of the average pressures 14 : 1. The average pressure is nearly inversely proportional to the distance, part of the difference probably being chargeable to the more nearly total reflection for the distant condition. It is implied, of course, that even for the latter condition the direct sound is large compared with that reaching the microphone by reflections.

So far the spectra discussed were those of normally modulated voices. Fig. 6 shows what happens to the average spectrum when a high, rather declamatory level is adopted. The higher level is relatively poorer in frequencies below 500 p.p.s., relatively richer between 500 and 4000 p.p.s., and above 4000 p.p.s. its spectrum is nearly the

same as for normal levels. The decrease at low frequencies is most pronounced in the band from 62 to 125 p.p.s., i.e. in the region of the voice fundamental frequency at normal levels. This leaves two possibilities. Either at the high speech level the fundamental frequency is the same as for normal, but its amplitude is relatively small; and the spacing of the overtones is the same as at normal levels. Or else—and more probably as indicated by the auditory pitch sense —passage to the high level is attended by an actual upward shifting of the fundamental frequency, and a correspondingly larger spacing

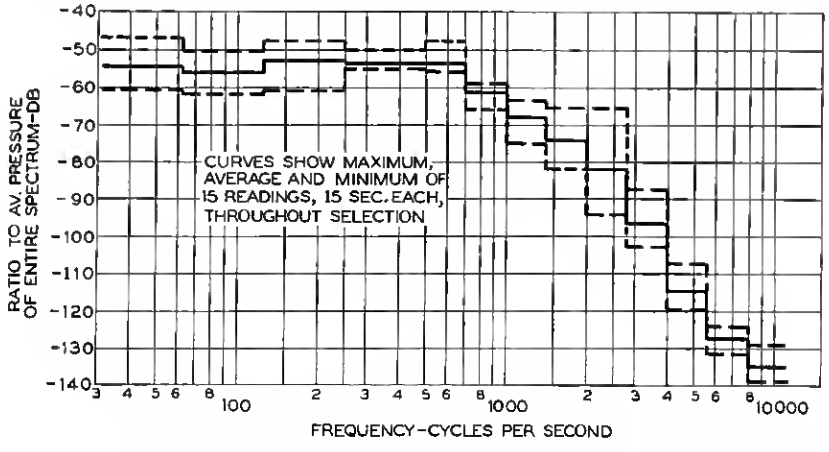

Fig. 7—Average pressures per frequency interval of 1 cycle per second—piano selection—Liszt's "Hungarian Rhapsody No. 2"—average total pressure = 3.5 bars.

of the overtones. In addition the amplitude of this new fundamental would be relatively lower than for the normal level. The average spectrum method is incapable of deciding between the two alternatives. Some oscillograms were made at normal and high speech levels, which clearly indicated that the second alternative is the correct one, or at least the prevalent one. For most vowel sounds, the fundamental frequency was found shifted from about 100 p.p.s. at normal to about 200 p.p.s. at the higher level. This result has an intimate bearing on the loss of naturalness encountered when speech originally picked up at normal voice levels is subsequently reproduced at much higher levels. A corresponding change, though in the opposite direction, probably takes place in going from normal to subnormal levels. Some evidence of this will be seen in the section on peak amplitudes.

The method of average pressure-frequency spectra is equally

applicable to sounds other than speech, provided they are sustained or can be repeated. Two illustrations are given. Fig. 7 is for a piano composition. Characteristic in comparison with speech is the relative smallness of high frequencies. Spectra of radically different types of compositions were found to be rather similar. The instrument, the room acoustics and the position of the microphone largely determine the picture. In this case, the microphone was about 7 ft. from the keyboard, and the reverberation time between 1.2 seconds at 100 p.p.s. and 1.0 second at 4000 p.p.s.

Fig. 8 gives the spectrum of street noise entering a laboratory

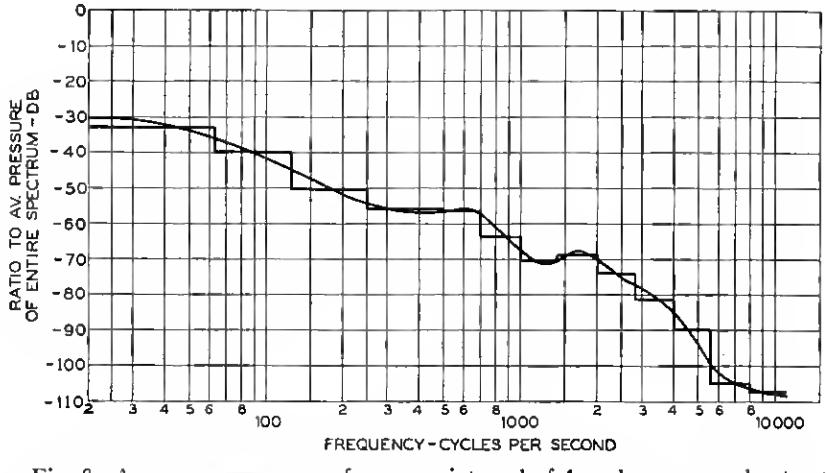

Fig. 8—Average pressures per frequency interval of 1 cycle per second—street noise—10th floor Bell Telephone Laboratories facing east—2.50–3.20 P.M., Oct. 10, 1928—average total pressure = 0.75 bar.

window, ten floors above the street. Street traffic and an elevated railway are the chief contributors, the measurements being taken during traffic peak hours. The relative poverty in high frequencies is even more pronounced than for the piano.

We shall now consider the type of apparatus intended primarily for measurements when the speakers are not acoustically controllable. Instead of averaging over a large number of words the measurement is essentially that of mean power in syllables. Usually it covers the whole spectrum rather than a particular frequency band. It has been widely used for controlling amplification in radio broadcasting, in phonograph and film recording of speech, and for rapid measurement and electrical control of speech levels in telephone conversations. A typical device of this sort is the "volume indicator" [9] shown in

[9] Essential features of this apparatus are shown by E. L. Nelson, U. S. Patent No. 1523827, filed 8/31/22.

Fig. 9, developed about ten years ago. Essentially it is a vacuum tube rectifier with a rapid action d.c. meter in the plate circuit. It is operated on a part of the characteristic such that the rectified plate current is roughly proportional to the square of the speech voltage. The rectifier is preceded by an amplifier of adjustable gain. For the speech level under measurement the gain is adjusted to such a value that the fluctuating meter deflections attain a prescribed

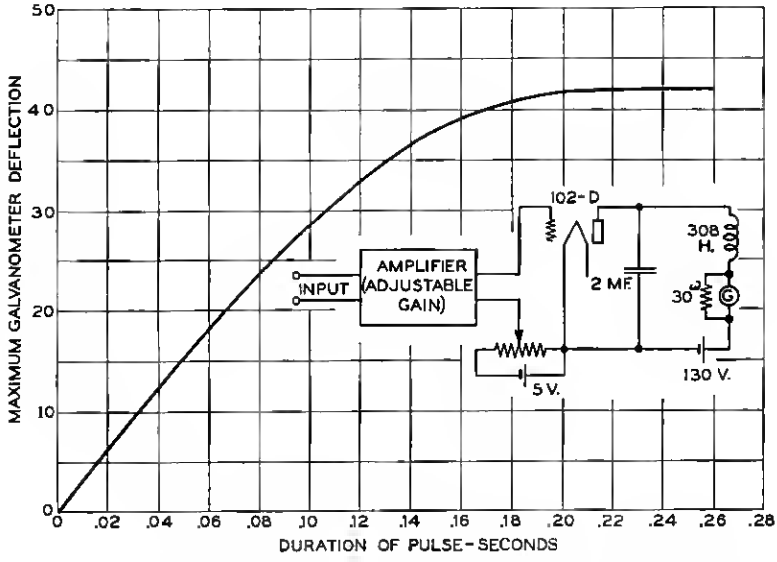

Fig. 9—Volume indicator—deflection as a function of a.c. pulse duration.

maximum value on the average of once in about three seconds. That value of gain, expressed in decibels with respect to a certain normal value of the gain, gives the volume indicator measure of the speech level. The meter, combined with the electric circuit, has a dynamic characteristic as shown in the curve, which gives the maximum deflection as a function of the duration of the a.-c. input. For inputs lasting more than about 0.18 second the maximum deflection remains the same. Since the average syllable duration is of the order of 0.2 second, it follows that the maximum deflection of the "volume indicator" meter is approximately proportioned to the mean power in the syllable. By comparison with oscillograms, or by equivalent methods, the volume indicator readings can be correlated by absolute quantities. Fig. 10 is an example showing the relation between speech level as measured with the volume indicator and the average

instantaneous amplitude of the speech waves on a certain laboratory telephone circuit of the commercial type.

Fig. 11 illustrates the type of data which can be expeditiously secured by means of the volume indicator. Here the levels for a

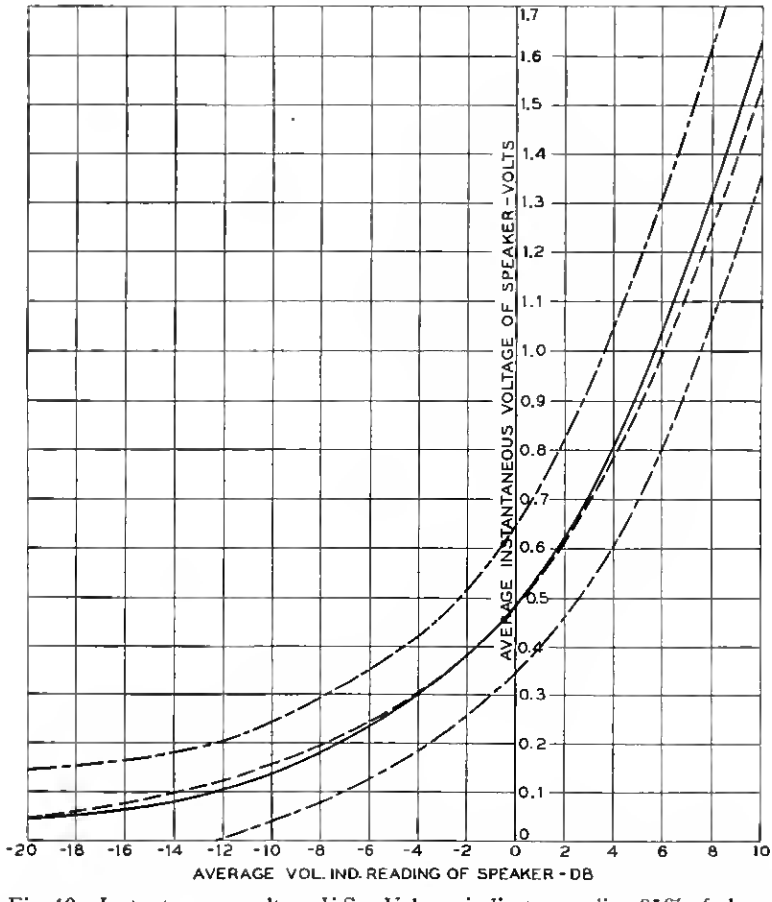

Fig. 10—Instantaneous voltage V.S. Volume indicator reading 85% of observed points included between outer dashed curves—ordinates of middle dashed curve are inversely proportional to gain of volume indicator.

large number of speakers and conversations (over a laboratory telephone circuit of the commercial type) are determined. To cover the same ground by means of oscillographic measurements would have been a decidedly formidable task.

Another device for the rapid measurement of speech levels is the "impulse meter," [10] shown in Fig. 12. This is essentially a peak

[10] D. Thierbach, *Zs. F. Techn. Physik*, No. 11, 1928.

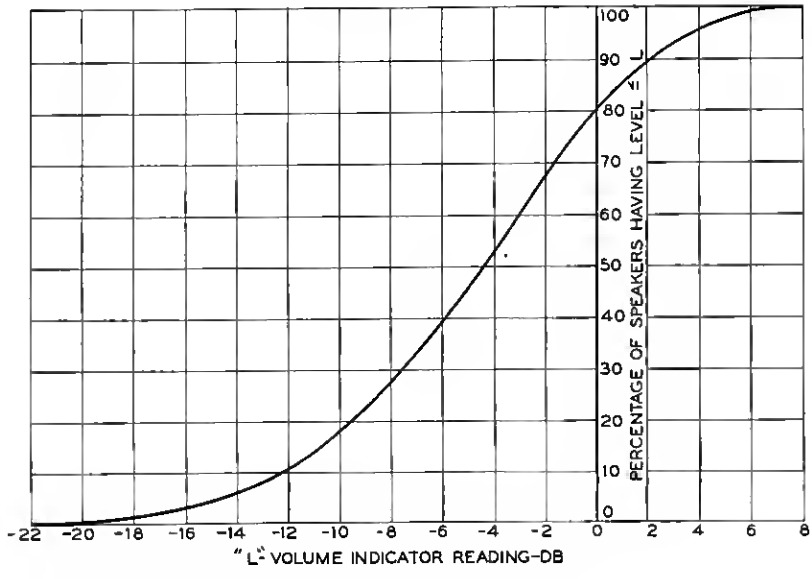

Fig. 11—Distribution of speaker's levels—87 men and 59 women.

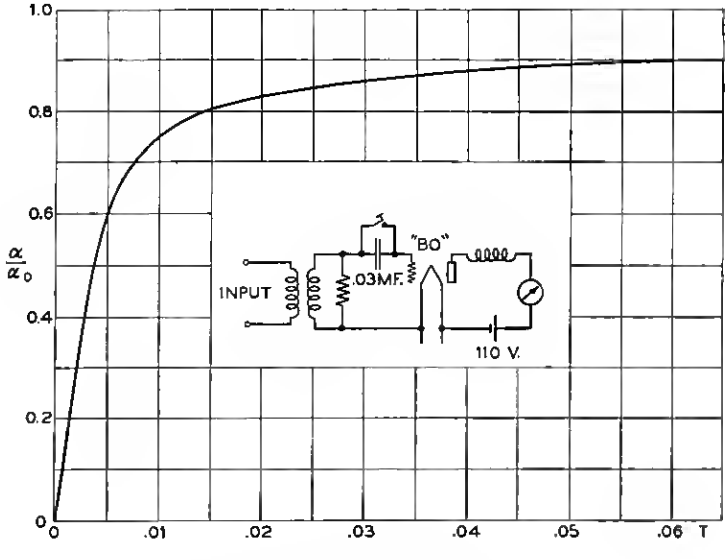

Fig. 12—Impulse meter.

T = Duration of a.c. impulse
α = Corresponding ultimate plate current
α_0 = Ultimate plate current for steadily applied a.c.

voltmeter. The curve shows the time rate at which the potential on the blocking condenser builds up. The time required for the galvanometer to reach its maximum deflection is determined by the dynamic characteristic of the meter and associated plate circuit, as well as by the time constant of the condenser charging circuit.

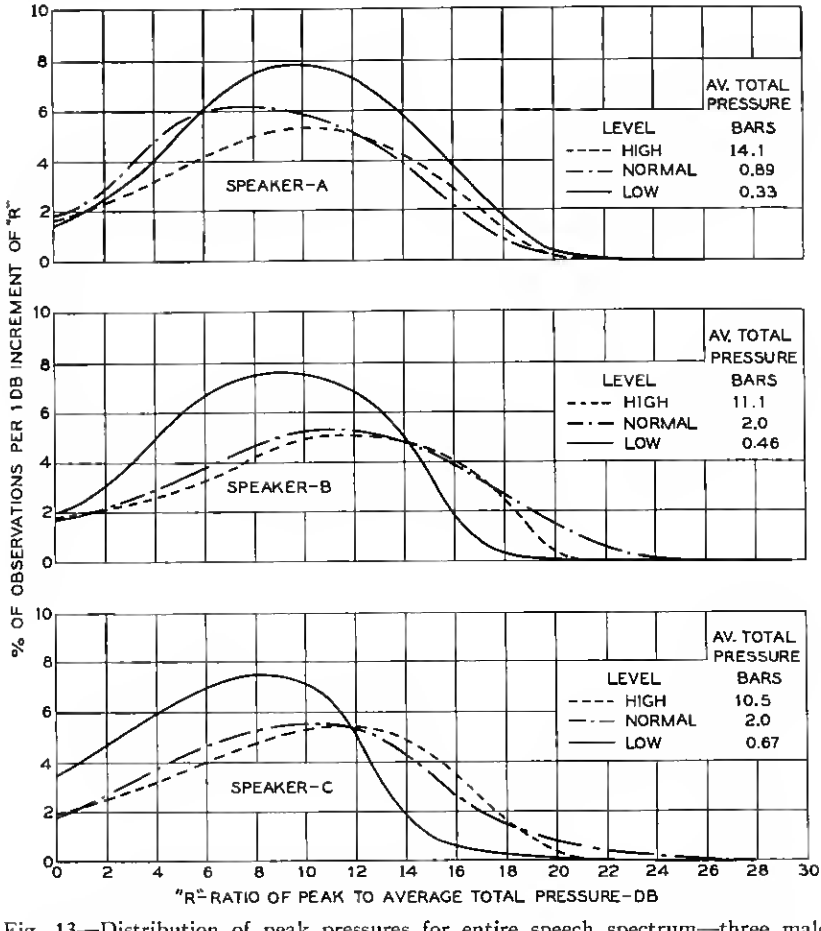

Fig. 13—Distribution of peak pressures for entire speech spectrum—three male voices—each peak is the maximum instantaneous pressure during 1/8 sec.

In using the volume indicator type of instrument in the telephone plant, limits are observed which have been determined by trial to give satisfactory operating results. The fact that it is possible to set such limits indicates a correlation between mean syllabic power and peak pressures. The overload capacity of apparatus is so important that

laboratory investigations of quantities affecting it merit more fundamental study than can be made with instruments of the volume indicator type. The quantities of interest are the instantaneous amplitudes of the wave peaks, the frequency of their occurrence and their distribution in the frequency spectrum. The apparatus used to obtain the data described below will be described in detail elsewhere. It is sufficient to state here that it is capable of directly (and automatically) registering the magnitudes and frequency of occurrence of the in-

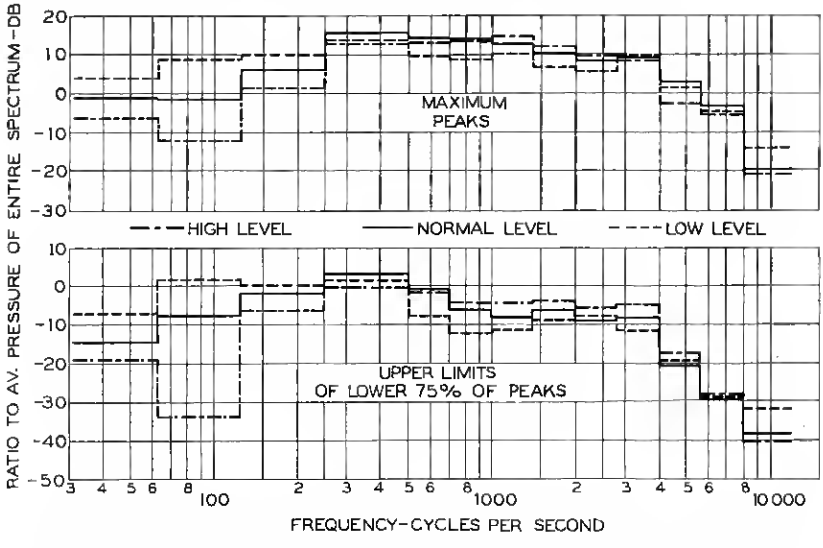

Fig. 14—Peak pressures of speech—composite, three male voices—each peak is the maximum instantaneous pressure during 1/8 sec.—average total pressure:
High (declamatory) level—8.7 bars;
Normal (conversational) level—1.6 bars;
Low (confidential) level—0.5 bar.

stantaneous peaks over a range of 60 db, corresponding to a power ratio of 1,000,000 : 1. The peak amplitudes were measured for the whole spectrum and also for restricted frequency bands selected by filters.

Fig. 13 and Fig. 14 show the results of some measurements with undistorted speech. Each individual observation gives the magnitude of the peak pressure in a 1/8 second interval. This is about as short an interval as one can use and still retain a high degree of probability that the individual measurements will give the maximum peak amplitudes in syllables. Otherwise, from the standpoint of the apparatus, the individual observations could be confined to much shorter time intervals, resulting in many more measurements for a

given length of speech. The ordinates give the ratios of the peak pressures to the average total speech pressure. In Fig. 13 the peaks are those of speech as a whole; in Fig. 14 the ordinates give the peak pressures in the several frequency bands. This is done for three widely different average levels spread over a range of 30 db. From

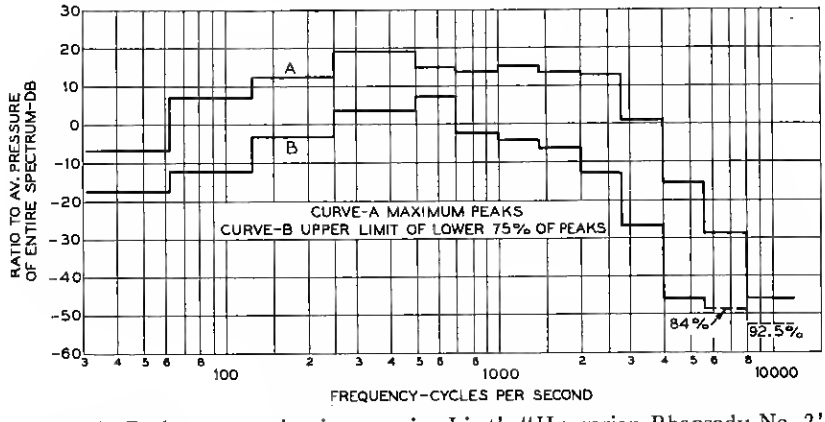

Fig. 15—Peak pressures in piano music—Liszt's "Hungarian Rhapsody No. 2" —each peak is the maximum instantaneous pressure during 1/8 sec.—average total pressure 4.0 bars.

the standpoint of providing apparatus overload capacity it is sometimes important to know not only the maximum values of all peaks (which might be uneconomical to provide for) but the upper limit of a certain percentage of all the peaks. The lower half of Fig. 14 illustrates this method of analyzing the peak data. It is interesting to note how much larger the transmission capacity must be to take care of the highest 25 per cent of the peaks.

Fig. 15 gives a similar picture of the peak amplitude distribution in a piano composition, for which the average amplitudes are given in Fig. 7. The microphone was about 7 ft. laterally from the center of the keyboard, and the measurements were made in a room having an average reverberation of about 1 second.

My thanks are due to Dr. H. K. Dunn and Mr. S. D. White of Bell Telephone Laboratories, who have obtained most of the data discussed in this paper.